Title
A possib le criterion for visual recognition thresholds

Author
Harris, James L

Publication Date
1959-11-01

Visibility Laboratory
University of California
Scripps Institution of Oceanography
San Diego 52, California

A POSSIBLE CRITERION FOR VISUAL RECOGNITION THRESHOLDS

James L. Harris

November 1959
Index Number NS 714-100

Bureau of Ships
Contract NObs-72092

SIO REFERENCE 59-65

Approved:

Seibert Q. Duntley

Seibert Q. Duntley, Director
Visibility Laboratory

Approved for Distribution:

Roger Revelle

Roger Revelle, Director
Scripps Institution of Oceanography

A POSSIBLE CRITERION FOR VISUAL RECOGNITION THRESHOLDS

James L. Harris

1.0 Introduction

Many military applications require the prediction of visual detection range and also the prediction of the range at which various levels of visual recognition can be performed. There is a considerable quantity of experimental visual psychophysics data which can be used for the purpose of predicting visual detection ranges. On the other hand there has been a deficiency in the type of experimental recognition data required to allow prediction of recognition ranges for situations of military interest.

The accumulation of sufficient psychophysical data to allow solution to all military visual recognition problems is a vast undertaking. An interim solution lies in brief psychophysical experiments, the results of which, though lacking generality, allow prediction for a specific situation.

Theoretical analyses can serve as a valuable guide to such experimentation. This report describes a theoretical analysis of the detection and recognition capability of an ideal mosaic detector. The relationship between detection and recognition for this idealized mosaic is used to hypothesize a criterion for the threshold of visual recognition. A brief psychophysics experiment was performed as a first test of this hypothesis. The degree of success and the limitations of the test are discussed.

2.0 Derivation of Equations

2.1.0 Detection

2.1.1 General

The model to be analyzed is as follows: The sensor is assumed to consist of an optical system of high quality which images the field of view on a mosaic of photosensitive elements. The object space field is assumed to be of uniform luminance. A target may or may not be present within this field of view. The target is describable by a spatial luminance map. The function of the sensor is to provide the best estimate as to whether or not a target is present. The ability of the sensor to perform this estimation is limited by the fact that each photosensitive element has an internal noise source assumed to be gaussian. It is assumed that the sensor system is capable of performing any desired operation on the receptor outputs.

2.1.2 Statistical Estimation

Consideration will first be given to the case in which the position of the target (if present) is fully known. It will also be assumed that the spatial luminance distribution of the target is fully known. For this case the analysis need include only those n receptors on which the image of the target falls.

Each receptor subtends a solid angle in object space. The target will be described by a set of luminances, B_1, B_2, $\cdots$, B_n, which are the target luminances averaged over the solid angle subtended by each receptor. The decision which the sensor system must make is whether the outputs of the receptors indicate the presence of a uniform luminance B_o due to the background alone, or a luminance map B_1, B_2, $\cdots$, B_n, due to the presence of the target.

If an ideal optical system is assumed then the flux incident on each detector will be

$$F = B \Omega A_L \tag{1}$$

where B is the luminance associated with each element, Ω is the solid angle subtended in object space by the receptor, and A_L is the area of the entrance pupil of the optical system. If it is further assumed that the photosensitive elements produce an output, i , which is linearly proportional to the incident flux, then

$$i = B \Omega A_L S \tag{2}$$

where S is the sensitivity of the receptor. The background luminance, B_o, then produces a receptor output, i_o, and the set of target luminances B_1, B_2, $\cdots$, B_n, produces a set of outputs i_1, i_2, $\cdots$, i_n.

It will be assumed that the target is present for a time T and that the starting time is known. Attention will first be directed to the output of a single receptor as a function of time. A digitalized approach will be taken by assuming that the output of the single receptor is sampled at very short time intervals, Δt. In the time interval T this results in a series of outputs, $x(t_1)$, $x(t_2)$, $\cdots$, $x(t_s)$. The first equation to be derived will be one which indicates the manner in which this set of outputs should be filtered in order to make the best estimate of the object space luminance associated with this receptor. The likelihood that a luminance, B , would result in an output x is

$$L = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\left(\frac{B \Omega A_L S - x}{2\sigma^2}\right)^2} \tag{3}$$

The likelihood that a luminance, B , would result in a set of outputs $x(t_1)$, $x(t_2)$, $\cdots$, $x(t_s)$ is

$$L = \prod_{i=1}^{\mathcal{A}} \frac{1}{\sqrt{2\pi}\,\sigma}\, e^{-\frac{(B \ln A_L S - X_i)^2}{2\sigma^2}} \tag{4}$$

or

$$L = \left[\frac{1}{\sqrt{2\pi}\,\sigma}\right]^{\mathcal{A}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{\mathcal{A}} (B \ln A_L S - X_i)^2} \tag{5}$$

The best estimate of the luminance is that value of B which will maximize L . This is the value which will minimize the summation. Let

$$y = \sum_{i=1}^{\mathcal{A}} (B \ln A_L S - X_i)^2 \tag{6}$$

then

$$\frac{dy}{dB} = \sum_{i=1}^{\mathcal{A}} 2(B \ln A_L S - X_i) \ln A_L S \tag{7}$$

Examining for a minimum

$$2 \sum_{i=1}^{\mathcal{A}} (B \ln A_L S - X_i) \ln A_L S = 0 \tag{8}$$

or (ignoring the trivial solution)

$$\sum_{i=1}^{\mathcal{A}} (B \ln A_L S - X_i) = 0 \tag{9}$$

or

$$A B \Omega A_L S = \sum_{i=1}^{A} X_i \tag{10}$$

or

$$B = \frac{\frac{1}{A} \sum_{i=1}^{A} X_i}{\Omega A_L S} \tag{11}$$

But since

$$\overline{X} = \frac{1}{A} \sum_{i=1}^{A} X_i \tag{12}$$

then

$$B = \frac{\overline{X}}{\Omega A_L S} \tag{13}$$

The best estimate of B is therefore determined by averaging the individual outputs. The first step in the processing is therefore to average the outputs of each receptor over the time period T. This results in a set of outputs for the n receptors of $\overline{X}_1, \overline{X}_2, \cdots, \overline{X}_n$. Each $\overline{X}$ will be a gaussian distribution of the form

$$f(\overline{x}) = \frac{1}{\sqrt{2\pi}\,\sigma \left(\frac{\Delta t}{T}\right)^{1/2}} e^{-\frac{(B\Omega A_L S - \overline{x})^2}{2\sigma^2 \left(\frac{\Delta t}{T}\right)}} \tag{14}$$

where B is the true luminance. The likelihood that a luminance, B, would result in an average $\overline{X}$ is as defined by Equation (14). The likelihood that a set of luminances $B_1, B_2, \cdots, B_n$ would result in a set of averages $\bar{X}_1, \bar{X}_2, \cdots, \bar{X}_n$ is

$$L = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\,\sigma \left(\frac{\Delta t}{T}\right)^{1/2}} e^{-\frac{(B_i \Omega A_L S - \bar{X}_i)^2}{2\sigma^2 \left(\frac{\Delta t}{T}\right)}} \tag{15}$$

or

$$L = \left[\frac{1}{\sqrt{2\pi}\,\sigma \left(\frac{\Delta t}{T}\right)^{1/2}}\right]^n e^{-\frac{\sum_{i=1}^{n} (B_i \Omega A_L S - \bar{X}_i)^2}{2\sigma^2 \left(\frac{\Delta t}{T}\right)}} \tag{16}$$

In order to make a decision the likelihood that the set of readings $\bar{X}_1, \bar{X}_2, \cdots, \bar{X}_n$ is due to a uniform luminance, B_o, must be computed. That is

$$L(B_o) = \left[\frac{1}{\sqrt{2\pi}\,\sigma \left(\frac{\Delta t}{T}\right)^{1/2}}\right]^n e^{-\sum_{i=1}^{n} \frac{(B_o \Omega A_L S - \bar{X}_i)^2}{2\sigma^2 \left(\frac{\Delta t}{T}\right)}} \tag{17}$$

The ratio of these two likelihoods forms a basis for making a decision as at the presence or absence of a target.

$$\frac{L}{L(B_o)} = e^{\frac{-\sum_{i=1}^{n} (B_i \Omega A_L S - \bar{X}_i)^2 + \sum_{i=1}^{n} (B_o \Omega A_L S - \bar{X}_i)^2}{2\sigma^2 \left(\frac{\Delta t}{T}\right)}} \tag{18}$$

A threshold could be established and a decision of target present made if the ratio exceeds some threshold value. However, since the numerator in the exponent is a monotonic function of the likelihood ratio, the function

$$y = -\sum_{i=1}^{n}\left(B_i \int\Omega A_L S - \bar{X}_i\right)^2 + \sum_{i=1}^{n}\left(B_0 \int\Omega A_L S - \bar{X}_i\right)^2 \quad (19)$$

can also be used as a decision function. If y is less than some value a target would be assumed to be present. The function y can be expanded and simplified as follows.

$$y = -\sum_{i=1}^{n}\left(B_i \int\Omega A_L S\right)^2 + \sum_{i=1}^{n}\left(B_0 \int\Omega A_L S\right) + 2\sum_{i=1}^{n} \bar{X}_i\left(B_i \int\Omega A_L S - B_0 \int\Omega A_L S\right) \quad (20)$$

For a known target and background the first two terms are constants while the last term describes the proper processing of the receptor outputs in order that the best decision be made.

2.1.3 Precision of Estimation

To determine how well the decision can be made it is necessary to determine the statistical properties of y as given in equation (20). It was previously noted that the first two terms in y are constants for a given target and background. The last term exhibits statistical fluctuations because of the noise inherent in $\bar{X}$. For example, if the target is actually present then $\bar{X}_i$ has a mean value of $B_i \int\Omega A_L S$ and a variance of $\sigma^2\left(\frac{\Delta t}{T}\right)$. The mean of the last term in equation (20) is therefore

$$\mu_1 = 2\sum_{i=1}^{n} B_i \Omega A_L S \left(B_i \Omega A_L S - B_o \Omega A_L S\right) \tag{21}$$

or

$$\mu_1 = 2\left(\Omega A_L S\right)^2 \sum_{i=1}^{n} B_i \left(B_i - B_o\right) \tag{22}$$

The variance of the last term is

$$\sigma_1^{\,2} = 2\sigma^2\left(\frac{\Delta t}{T}\right)\sum_{i=1}^{n} \left(B_i - B_o\right)^2 \left(\Omega A_L S\right)^2 \tag{23}$$

The mean of equation (20) is therefore

$$\bar{y} = \left(\Omega A_L S\right)^2 \left[-\sum_{i=1}^{n} B_i^2 + \sum_{i=1}^{n} B_o^2 + 2\sum_{i=1}^{n} B_i^2 - 2\sum B_i B_o\right] \tag{24}$$

or

$$\bar{y} = \left(\Omega A_L S\right)^2 \left[\sum_{i=1}^{n} B_i^{\,2} - 2\sum_{i=1}^{n} B_i B_o + \sum_{i=1}^{n} B_o^{\,2}\right] \tag{25}$$

or

$$\bar{y} = (\Omega A_L S)^2 \left[\sum_{i=1}^{n} (B_i - B_o)^2\right] \tag{26}$$

If the target is not present then

$$\bar{y} = -(\Omega A_L S)^2 \left[\sum_{i=1}^{n} (B_i - B_o)^2\right] \tag{27}$$

and the variance is the same as that given in equation (23). The statistics of the decision can be most easily visualized by making the substitution

$$Z = \frac{y}{\sigma \sqrt{2\left(\frac{\Delta t}{T}\right) \sum_{i=1}^{n} (B_i - B_o)^2 (\Omega A_L S)^2}} \tag{28}$$

Then

$$\sigma_Z = 1 \tag{29}$$

$$\bar{Z} = \pm \frac{\left[\sum_{i=1}^{n} (B_i - B_o)^2\right]^{1/2} \Omega A_L S}{\sigma \sqrt{2\left(\frac{\Delta t}{T}\right)}} \tag{30}$$

The sign of $\overline{Z}$ is positive if a target is present and negative if a target is not present.

Equation (30) can be used to determine the detection and false alarm probabilities associated with the mosaic sensor. For the purpose of this report however it is sufficient to point out that the term

$$\alpha = \left[\sum_{i=1}^{n} \left(B_i - B_o\right)^2\right]^{1/2} \tag{31}$$

specifies the detectability of a target. Equation (31) can be rewritten by defining the "point" contrast as

$$C_i = \frac{B_i - B_o}{B_o} \tag{32}$$

and therefore

$$\alpha = B_o\left[\sum_{i=1}^{n} C_i^{\,2}\right]^{1/2} \tag{33}$$

2.1.4 Interpretation of Detectability Criterion

Equation (33) indicates the manner in which the detectability is determined for a target having internal contrast variation. Note that for the special case of C_i a constant

$$\alpha = B_o\, C\, n^{1/2} \tag{34}$$

Here contrast is the measure of detectability. Equation (33) implies that if a target has both negative and positive contrast components they will both contribute as described by the quadratic summation of the components. Because of the quadratic addition, target components of high contrast will tend to predominate in determining detectability. For example assume a target having two equal areas, one of which has twice the contrast of the other. The existence of the area of lower contrast will alter α by only about 11%.

2.2.0 Recognition

2.2.1 General

The same simple model can be extended to the case of recognition. It will be assumed that the mosaic sensor system is required to make a decision as to whether the object detected is Target A or B.

2.2.2 Statistical Estimation

The derivation is identical to that previously given except that in place of making a decision between B_o and $B_1, B_2, \cdots, B_n$, it is now required that a decision be made between $B_{A_1}, B_{A_2}, \cdots, B_{A_n}$ and $B_{B_1}, B_{B_2}, \cdots, B_{B_n}$. This can be accomplished by direct substitution in equation (20) to give

$$y = -\sum_{i=1}^{n} \left(B_{A_i} \Omega A_L S\right)^2 + \sum_{i=1}^{n} \left(B_{B_i} \Omega A_L S\right)^2 + 2\sum_{i=1}^{n} \bar{x}_i \left(B_{A_i} \Omega A_L S - B_{B_i} \Omega A_L S\right) \quad (35)$$

It is worthwhile to note that the last term in y is the difference of two correlation-type processes. That is $\sum_{i=1}^{n} \overline{X}_i \left(B_{A_i} \mathcal{N} A_L S\right)$ is proportional to the correlation of the outputs obtained from the sensor and the mean outputs which would be obtained if Target A was present. In a similar manner $\sum_{i=1}^{n} \overline{X}_i \left(B_{B_i} \mathcal{N} A_L S\right)$ is proportional to the correlation of the outputs obtained and the mean outputs which would be obtained if Target B was present.

2.2.3 Precision of the Estimate

For the case of recognition equation (31) becomes

$$\alpha = \left[\sum_{i=1}^{n} \left(B_{A_i} - B_{B_i}\right)^2\right]^{1/2} \tag{36}$$

The recognizability of two targets is determined by the quadratic content of the "difference image".

3.0 Recognition Hypothesis

The recognition hypothesis is generated by comparison of Equations (31) and (36). It may be noted that both equations indicate that performance is dictated by the quadratic content of a "difference image". For the case of detection, it is a point by point difference between the target and the background. For the case of recognition it is a point by point difference between Target A and Target B.

Hypothesis: The capability of making a decision as to which of two possible visual targets is present (recognition) is equal to the capability of making a decision as to the presence or absence of the "difference image" (detection).

This hypothesis can be examined for two special cases of interest. First, suppose that two objects have identical shapes and luminances except for some unique protuberance on one of them. It seems reasonable to suppose that this protuberance, which is the difference image must be detectable before discrimination can be performed.

Second, consider the more general case of two targets of arbitrary shape and luminance pattern. It would seem intuitively reasonable that the greater the correlation between the two targets, the shorter the recognition range. Equation (36) can be rewritten in terms of cross-correlation between the two targets as follows:

$$\alpha = \left[\sum_{i=1}^{n} \left(B_{A_i}^{2} - 2 B_{A_i} B_{B_i} + B_{B_i}^{2} \right) \right]^{1/2} \qquad (37)$$

or

$$\alpha = \left[\sum_{i=1}^{n} \left(B_{A_i}^{2} + B_{B_i}^{2} \right) - 2 \sum_{i=1}^{n} B_{A_i} B_{B_i} \right]^{1/2} \tag{38}$$

or

$$\alpha = \sqrt{\sum_{i=1}^{n} B_{A_i}^{2} \sum B_{i}^{2}} \left[\frac{\sum_{i=1}^{n} \left(B_{A_i}^{2} + B_{B_i}^{2} \right)}{\sqrt{\sum_{i=1}^{n} B_{A_i}^{2} \sum B_{B_i}^{2}}} - 2 \frac{\sum_{i=1}^{n} B_{A_i} B_{B_i}}{\sqrt{\sum_{i=1}^{n} B_{A_i}^{2} \sum B_{B_i}^{2}}} \right]^{1/2} \tag{39}$$

But

$$\rho_{AB} = \frac{\sum_{i=1}^{n} B_{A_i} B_{B_i}}{\sqrt{\sum_{i=1}^{n} B_{A_i}^{2} \sum B_{B_i}^{2}}} \tag{40}$$

where ρ_{AB} is the cross-correlation between Targets A and B. Therefore,

$$\alpha = \sqrt{\sum_{i=1}^{n} B_{A_i}^{2} \sum B_{B_i}^{2}} \left[\frac{\sum_{i=1}^{n} \left(B_{A_i}^{2} + B_{B_i}^{2} \right)}{\sqrt{\sum_{i=1}^{n} B_{A_i}^{2} \sum_{i=1}^{n} B_{B_i}^{2}}} - 2 \rho_{AB} \right]^{1/2} \tag{41}$$

For the special case of $\sum_{i=1}^{n} B_{A_i}^{2} = \sum_{i=1}^{n} B_{B_i}^{2}$

$$\alpha = 2 \sum_{i=1}^{n} B_{A_i}^{2} \left[1 - \rho_{AB} \right]^{1/2} \tag{42}$$

Equation (41) shows the relationship between the cross-correlation, ρ_{AB}, and recognition capability.

4.0 Test of the Hypothesis

Figure 1 (a, b, c, d) shows four silhouette type targets used in the psychophysics experiment.* The dimensions are shown in angular subtense in minutes of arc from the point of observation. The test was forced choice in time. A buzzer sounded at 2-second intervals marking off a total of four intervals. During one of these intervals, chosen at random, one of the four targets was projected. The observer was required to specify the interval and the target number (detection and recognition). The data from the experiment, after correction for chance, is shown in Figure 2. The curves are the detection and recognition probabilities averaged over all targets.

In order to check the hypothesis of Section 3, the "difference images" must first be determined. They are shown in Figure 1 (e through j). A thorough analysis should include the summative properties of the eye. It may be noted, however, that to a first approximation all of the images are spread over a comparable spatial area. Therefore, for a cursory test the relative areas of the targets will be compared.

The areas of Targets (a) through (d) are 338 square minutes. The average area of Targets (e) through (j) is 197 square minutes. The ratio of these two areas is 1.715. It is therefore assumed that the contrast required for recognition would be 1.715 times that required for detection. Figure 1 shows a calculated recognition curve which is the detection curve displaced in contrast by a factor of 1.715. The agreement with the experimental curve is remarkably good.

---

* The experiment was designed, executed and analyzed by Dr. John H. Taylor of this laboratory. Dr. Taylor is however not responsible for the use to which this data has been subjected or to the conclusions drawn.

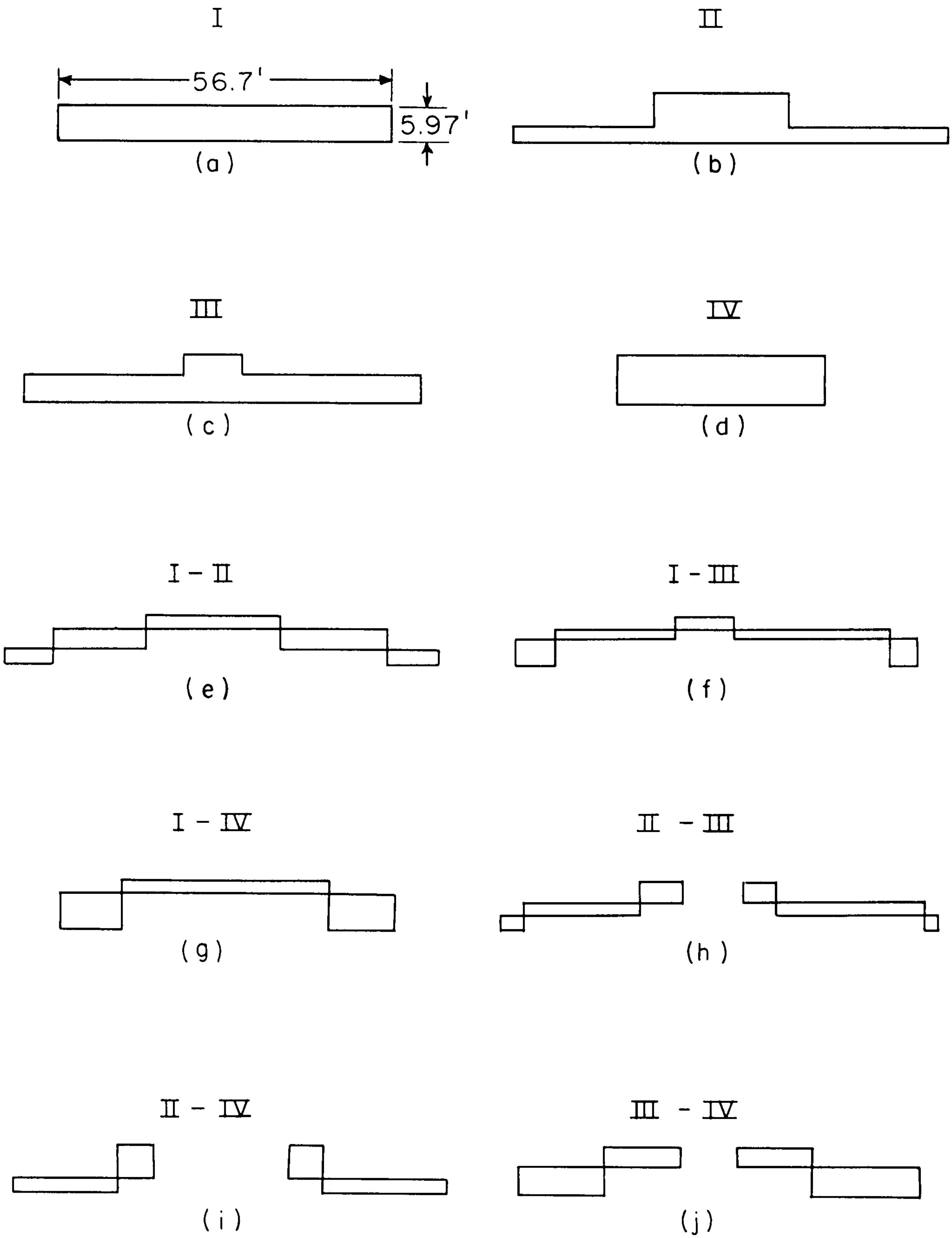

Figure 1 Experimental Targets and Their Difference Images

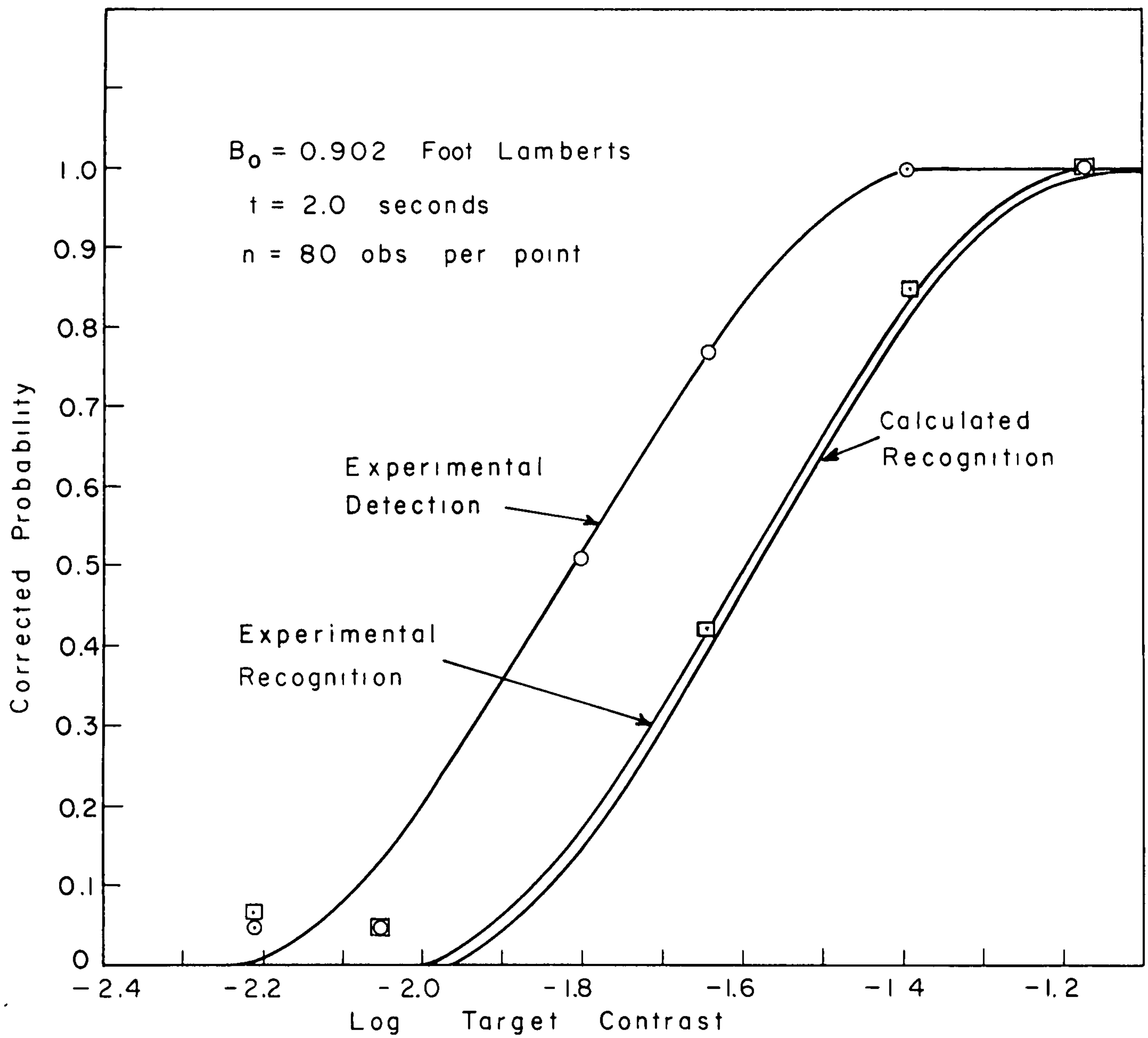

Figure 2 Comparison of Calculated and Experimental Recognition Performance

5.0 Conclusions

The derivation of the recognition equations for the ideal mosaic sensor produced the following significant results:

5.1 Concept of Difference Image

First it indicated that the ability to discriminate between two targets is determined by the quadratic content of the "difference image". This seems intuitively clear. That is if two targets differ only by a protuberance on one, the protuberance must be detected in order to perform discrimination. Secondly, it was shown that for the idealized mosaic sensor, the discrimination between two objects bears a definable relationship to the cross-correlation between the two objects.

5.2 Test of the Recognition Hypothesis

The psychophysical experiment seems well explained by the "difference image" hypothesis. As a matter of fact, it seems too well explained considering the crude manner in which the detectability of the ten different targets were compared (area consideration only). The results however were encouraging and further comparison of theoretical and experimental results should be performed.

5.3 Calculation of Visual Recognition Ranges

Finally, on the basis of the evidence presented here, and until such time as more refined theoretical and experimental work can be performed, it seems reasonable, when numerical recognition ranges are required, to assume that discrimination between two targets is equivalent to detection of the difference image.